

SYSTEM AND METHOD FOR JITTER COMPENSATION IN DATA TRANSFERS

Ans a1

FIELD OF THE INVENTION

5 The invention relates generally to data communication in computer systems, and more particularly to a system and method implemented in connection with serial data packet protocols which re-time received frames to an asynchronous transmit clock wherein overflow or underflow conditions are detected in
10 an elastic buffer and corrected.

BACKGROUND OF THE INVENTION

15 Individual computer systems can be combined to form networks. There are many different types of networks, including local area networks (LANs,) wide area networks (WANs,) storage area networks (SANs) and many others. Networks are typically characterized by several characteristics, including their protocols, their
20 architectures and their topologies.

One of the network topologies which is commonly used is a ring, or loop. This topology may be implemented, for example, in a Fibre Channel Arbitrated Loop (FC-AL.) Figure 1 shows a
25 network which implements a loop topology.

In a network having a loop topology, data which needs to be transmitted from one device (e.g., a computer) in the network to another must be passed from the first device to the
30 next device in the loop, which passes the data to the next device, and so on, until the data reaches the destination device. The data is typically formatted into a plurality of frames (or packets of some sort) which carry the payload (the

Gray Cary\AU\4036099.2

data) and also contain overhead information which, for example, identifies the destination device to which the data is directed. If this device is the destination of the data, the receiving device accepts the data and uses it as needed.

- 5 If this device is not the destination device, the frame is transmitted to the next device in the network.

The devices in the network typically read the incoming frame data by reconstructing (from the data itself) the clock signal of the device from which the data is received. If it is determined that the data should be passed on to the next device in the loop, the data is typically retransmitted according to a clock signal which is generated within the device which is retransmitting the information. A second clock signal is used to transmit the data because, if the received clock signal were used the amount of jitter added by each node would be cumulative. As a result, the jitter could increase enough to corrupt the data.

Because the recovered receive clock and the transmit clock may have slightly different frequencies, the amount of data received in a given period may not be equal to the amount of data transmitted in that same period. As a result, the buffer in which the received data is stored until it is retransmitted may exceed its capacity (an overflow condition,) or there may not be sufficient data in the buffer to be retransmitted (an underflow condition.) This problem can be compounded each time the frames pass through one of the devices in the loop. Unless there is some means to compensate for this problem, corruption of the data may occur. A mechanism is therefore required to account for the possible differences between the clock signals in the different devices in the network.

One protocol which is used in some networks is the Fibre Channel Arbitrated Loop (FC-AL) protocol. In this protocol, several fill words are transmitted between each of the frames of data. The fill words are simply that -- non-useful or redundant words which fill in the space between frames. If the differences in the clock signals cause the frames to be transmitted more slowly than they are received (a potential overflow,) one of the fill words can be deleted from the retransmitted data in order to allow the transmission to "catch up" with the received data. If the frames are retransmitted more quickly than the data is received (a potential underflow,) a fill word can be added between frames to effectively slow down the retransmission and allow the received data to catch up.

SUMMARY

One or more of the problems outlined above may be solved by the various embodiments of the invention which, broadly speaking, comprises a system and method for inserting or deleting fill words between frames of data to compensate for frequency differences between a recovered receive clock and an internal transmit clock.

In one embodiment, the system comprises a circularly accessed buffer coupled to read and write logic. The read and write logic access data words at locations within the circular buffer as indicated by respective read and write pointers. The system further comprises control logic which compares the pointers to determine whether the buffer is approaching an underflow or overflow condition and adds or deletes fill words between frames of data to compensate for the impending underflow or overflow condition. Fill words are added or deleting by manipulating the read or write pointer to re-read or overwrite fill words when necessary. In one embodiment, the system includes fill word logic which is configured to add a fill word bit to each received word and to set or clear the fill word bit to indicate whether or not the corresponding word is a fill word. In one embodiment, the system compares read and write pointers only when one of the pointers has a value of 0 in order to reduce the number of times they are compared and thus to simplify the comparison logic.

In one embodiment, a method comprises writing words to locations in a circular buffer indicated by a write pointer, and reading words out of locations in the buffer indicated by a read pointer. The method further comprises periodically comparing the pointers to identify an approaching overflow or

underflow condition and manipulating the read or write pointers in response to the detected overflow or underflow condition. If an approaching overflow condition is detected, the write pointer is manipulated to overwrite a fillword. If
5 an approaching underflow condition is detected, the read pointer is manipulated to re-read a fillword. In one embodiment, the method further comprises adding a fill word bit to each word stored in the buffer to facilitate identification of fill words. In one embodiment, the pointers
10 are compared only when one of the pointers has a value of 0. In one embodiment, the fill words may be deleted with higher priority if the spacing between the read and write pointers exceeds a predetermined maximum threshold.

BRIEF DESCRIPTION OF THE DRAWINGS

Other objects and advantages of the invention may become
5 apparent upon reading the following detailed description and
upon reference to the accompanying drawings in which:

Fig. 1 is a block diagram illustrating a series of
devices which are connected in a ring or loop configuration.

10

Figs. 2a and 2b are block diagrams illustrating the
routing of data by a device (a) when the device is the
destination of the data and (b) when the device is not the
destination and passes the data through to a subsequent
15 device.

Fig. 3 is a functional block diagram of the elastic
buffer and associated logic in one embodiment.

20

Figs. 4a and 4b are very simplified state diagrams
illustrating the operation of these state machines based upon
the spacing of the read and write pointers in one embodiment.

25

While the invention is subject to various modifications
and alternative forms, specific embodiments thereof are shown
by way of example in the drawings and the accompanying
detailed description. It should be understood, however, that
the drawings and detailed description are not intended to
limit the invention to the particular embodiment which is
30 described. This disclosure is instead intended to cover all
modifications, equivalents and alternatives falling within the
spirit and scope of the present invention as defined by the
appended claims.

ATTY DOCKET NO.
CROSS1410-1

PATENT APPLICATION

- 7 -

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

A preferred embodiment of the invention is described below. It should be noted that this embodiment and other 5 embodiments described below are exemplary and are intended to be illustrative of the invention rather than limiting.

Broadly speaking, the invention comprises a system and method for using pointers to indicate read and write locations 10 in a clocked circular buffer and comparing the pointers to determine whether fill words between frames should be added or deleted to prevent overflow or underflow of data from the buffer. If the write pointer leads the read pointer by more than a predetermined amount (approaching overflow,) the write 15 pointer pauses and overwrites a fillword (effectively deleting a fillword,) allowing the read pointer to catch up to the write pointer. If the write pointer leads the read pointer by less than a predetermined amount (approaching underflow,) the read pointer pauses, re-reading the same fill word 20 (effectively adding another fill word to the transmitted data stream) to allow the write pointer to restore the desired lead.

One embodiment of the present system is implemented in a 25 Fibre Channel network. "Fibre Channel" refers to an data communication technology, as well as a set of standards being developed by the American National Standards Institute (ANSI) to define the technology. Fibre Channel supports both shared media and dedicated, or switched, media. Fibre Channel can be 30 used to provide means for data transfer in many different systems, supporting workstation clusters, switched LANs, SANS (storage area networks) and the like.

Despite its name, "Fibre Channel" technology is not strictly a channel-oriented technology, nor does it represent a network topology. Fibre Channel allows devices to be interconnected in a more generalized scheme. In this scheme, 5 devices may be connected by Fibre Channel links in configurations which include point-to-point configurations, switched fabric configurations and loop configurations. Point-to-point configurations are simply direct connections between two devices. "Switched fabric" configurations use 10 networks formed by one or more interconnected Fibre Channel links and switches. One device in the network can be connected to any of the other devices in the network through the "fabric" formed by the links and switches. Loop configurations, such as Fibre Channel - Arbitrated Loop, or 15 FC-AL, allow data to be communicated between any two of the devices in the network, but the data may have to pass through one or more of the other devices in the network.

Referring again to Fig. 1, the devices coupled to an FC- 20 AL network are connected in series, with each device being coupled to receive data from a preceding device and to transmit data to a succeeding device. Referring to Figs. 2a and 2b, each of the devices 10 includes a switch 12 and an elastic buffer 14. Switch 12 is configured to route the 25 frame, whether that frame is destined for corresponding device 10 or a subsequent device in the loop. If the data is destined for current device 10, switch 12 routes the data into the device for processing as shown in Fig. 2a. In this case, the device forwards canned fillwords to the next device in the 30 loop. If the data is not destined for current device 10, but is instead destined for a subsequent device, switch 12 routes the data to pass through the current device as shown in Fig. 2b. If the data is passed through, it is routed to elastic

Gray Cary\AU\4036099.2

buffer 14, which is configured to transmit it to the next device in the loop.

Referring to Fig. 3, a functional block diagram of the elastic buffer and associated logic of one embodiment is shown. In this embodiment, data which is passed through a device is written by write logic 22 to buffer storage 26. Write logic 22 is clocked by the clock signal recovered from the received data. The locations to which the data is written are determined by the value of the associated write pointer 23. Write logic 22 advances write pointer 23 as each storage location is written, so that each successive half-word received from the preceding device in the loop is stored in a successive storage location in the buffer. Buffer storage 26 is configured as a circular buffer. That is, after the write pointer (or read pointer) reaches the last storage location in the buffer, it "wraps around" and points to the first storage location again.

As indicated above, one half-word is written to each storage location in this embodiment. This is a result of using 16-bit wide RAM to store data which, as defined by the Fibre Channel specification, comprises 32-bit words. It should be noted that, in other embodiments, different word lengths and storage locations of different widths may be used, so that each storage location may store full words, portions of words or multiple words. These variations are contemplated to be within the scope of this disclosure. It should be noted that, for the purposes of this disclosure, the term "word" is used to describe a portion of the data which is stored in one of the storage elements of the buffer, and is not limited to words of a particular bit-length.

In one embodiment, buffer storage 26 is implemented using dual port synchronous RAM as storage elements. Each element is 16 bits wide and can store half of a 32-bit Fibre Channel data word. The protocol defined by the Fibre Channel specification also includes a K flag in each word. The present system also includes a fill word flag in each word. This embodiment therefore includes registers to store two additional bits (the fill word bit and the K flag bit) for each of the 16-bit half-words in the RAM storage elements.

Since each storage element only stores a half-word in this embodiment, the fill word bit and K flag bit for each word are stored with the corresponding high half-word.

Read logic 24 reads the data out of buffer storage 26 for transmission to the succeeding device in the loop. Read logic 24 is clocked by an internally generated transmit clock which, although it is intended to operate at the same frequency as the recovered clock signal, is independent of this signal and may therefore vary from it to some extent. Read logic 24 is configured to retrieve the data from the locations indicated by read pointer 25. In operation, read pointer 25 should point to a storage location which differs from the target of the write pointer by an amount which is implementation dependent. The Fibre Channel specification requires that the latency of data which is passed through a device to be a maximum of six word-times. The amount by which the write pointer leads the read pointer adds to the latency, so the spacing of the pointers should be minimized, but should still allow for the addition or deletion of words without corrupting data. In one embodiment, buffer storage 26 has eight storage elements, and the ideal spacing for the pointers is three or four storage locations.

The present system also includes control logic 28, which is coupled to both read logic 24 and write logic 22. Control logic 28 is configured to periodically access read pointer 25 and write pointer 23 to determine the spacing of the storage locations to which they point. If the indicated read and write locations are optimally spaced, the read and write logic proceed without interruption or interference from the control logic. If the indicated read and write locations are too close together (i.e., the write pointer leads the read pointer by less than a predetermined amount,) control logic 28 is configured to cause the read logic to re-read one of the storage locations before continuing and reading the succeeding storage locations. This causes the spacing between the pointers to increase. If, on the other hand, the indicated read and write locations are too far apart (i.e., the write pointer leads the read pointer by more than a predetermined amount,) control logic 28 is configured to cause the write logic to overwrite the same storage location. This causes the spacing between the pointers to decrease.

It should be noted that control logic 28 does not cause read logic 24 to re-read or skip words that contain any of the useful information being transmitted from the source computer to the destination computer. Doing so would corrupt the data. Only fill words, which contain redundant information, are allowed to be skipped or re-read. Thus, fill words are effectively added or deleted between frames when the data is retransmitted by the device.

In the Fibre Channel specification, specific bit sequences are defined as fill words. Fill words are normally detected by scanning each received word and determining which of the words match the sequences defined in the Fibre Channel

Gray Cary\AU\4036099.2

specification. To facilitate identification of fill words in the present system, this function is performed by fill word logic 30. Fill word logic 30 is positioned upstream from write logic 22 and is configured to add a fill word flag bit 5 to each of the words which is written to buffer storage 26. This bit is set for the fill words and cleared for all other words. After the fill word bit is added to each of the words, fill words can be identified by checking this bit, rather than scanning the entire word to determine whether it matches one 10 of the predefined fill word sequences.

Control logic 28 uses two state machines to control the addition or deletion of fillwords between frames. Referring to Figs. 4a and b, two state diagrams illustrating the operation 15 of these state machines are shown. In one embodiment, the system is only periodically checked to determine whether the read and write pointers are properly spaced. Because the difference between the recovered receive clock and the internal transmit clock is expected to be minimal, there is no 20 need to check the spacing of the pointers every time the system writes to or reads from one of the storage locations. The timing of the periodic checks is discussed in more detail below.

Before describing the operation of the state machines in control logic 28, it should be noted that the spacing between 25 pointers in this embodiment is determined by subtracting the location of the read pointer from the location of the write pointer. In other embodiments, the spacing can be determined 30 by subtracting the location of the write pointer from the location of the read pointer. When either state machine is considered, one of these cases will involve a comparison of the pointer spacing to a minimum threshold, while the other

Gray Cary\AU\4036099.2

case will involve a comparison of the pointer spacing to a maximum threshold. For instance, if a circularly accessed buffer has 8 storage elements, and the write pointer should lead the read pointer by no more than 5 storage locations, the
5 spacing determined by subtracting the read pointer from the write pointer is compared to a maximum value of 5. If, on the other hand, the spacing is determined by subtracting the write pointer from the read pointer, the spacing is compared to a minimum value of 3 (8-5).

10

As noted above, this embodiment determines pointer spacing by subtracting the location of the read pointer from the location of the write pointer. Referring to Fig. 4a, when the spacing between pointers is checked, the addition state machine (i.e., the state machine that controls the addition of fill words) compares the pointer spacing to a predetermined minimum value. If the spacing is greater than the predetermined minimum value, it is not necessary to add a fill word, and the state machine remains in the idle state. If, on the other hand the spacing is less than the predetermined minimum value, a fill word should be added, and the state machine moves to state S1. When the state machine moves to state S1, the system examines the words which are being read from buffer storage 26 by read logic 24. As explained above,
20 only fill words are re-read, so the control logic takes no further action until a fill word is detected. When a fill word is detected, the state machine moves to state S2, and a fill word is added to the stream of transmitted data by reading the fill word twice (the read pointer is not
25 incremented after the first read.)
30

Referring to figure 4b, the operation of the deletion state machine is similar to that of the addition state

machine, except that the pointer spacing is compared to a predetermined maximum value, and a fill word is deleted (overwritten) rather than being re-read. The deletion of the fillword occurs according to the high- and low- priority
5 conditions described in more detail below.

In an embodiment using eight storage locations, the write pointer initially leads the read pointer by 3 or 4 storage locations. If the spacing is determined by subtracting the
10 read pointer from the write pointer, there is an initial spacing of four or five storage locations. This spacing is considered to be safe, and no action needs to be taken if this spacing is detected. If the spacing is 0, 1 or 7, the
15 pointers are too far apart and an error is considered to have occurred. If the spacing is 2 or 3, a fill word should be added and, if the spacing is 6, a fillword should be deleted.

It should be noted that, in this embodiment, spacings of 4 or 5 are both quiescent because each storage location stores
20 a half word, while full words are added or deleted. If only one spacing was quiescent, then each time a word was added or deleted, the pointers would end up with spacing that would indicate an opposing action. In other words, every time a word was added, the resulting spacing would indicate that a
25 word should be deleted, and vice versa. Consequently, the system would constantly be adding and deleting words.

In one embodiment, the system may take into account the urgency of the need to add or delete a fill word. It is
30 helpful to consider the use of fill words between frames in a Fibre Channel system. According to the Fibre Channel specification, the number of fill words between frames should be a minimum of 6 when the frames are originated and should
Gray Cary\AU\4036099.2

always be at least 2 when received. Consequently, the control logic is configured so that, if it has detected the need to delete a fill word, it will count the number of fill words between frames before deleting a fill word. If deleting one
5 of the existing fill words would result in less than 4 fill words between the frames, the control logic may delay the deletion unless it is a high priority action. In other words,
10 if the data may otherwise be corrupted, it may be necessary to immediately delete a fill word. If the deletion is a low priority, however, the control logic may wait until the next series of fill words is detected before deleting a fill word.
15 (In the system described above, where pointer spacing of 2 or 3 storage locations indicate the need to delete a fill word, the spacing of 2 may be a high priority deletion, whereas a spacing of 3 may be a low priority deletion.)

In one embodiment, control logic 28 performs an asynchronous comparison between the read and write pointers and has an error of plus or minus one storage location. Even a slight variation between the receive and transmit clocks can cause this circuit to dither between two values. This may cause the control logic to add a fill word and then almost immediately delete a fill word, or vice versa. This embodiment therefore implements a form of hysteresis by including logic to suppress any addition or deletion of fill words unless the pointer spacing is stable for at least half of a clock cycle.

This embodiment also implements a dead time to the add and delete state machines. Because the read and write pointers are only checked every eighth clock cycle, it takes sometime before the spacing between the pointers can be updated to reflect the results of adding or deleting a word.
Gray Cary\AU\4036099.2

Consequently, after a fill word is added or deleted, the state machines are disabled for about 15 clock cycles so that incorrect spacing information is not generated.

- 5 In order to simplify the calculation of the spacing between the pointers, the spacing is determined when one of the pointers (more specifically, the pointer which is subtracted from the other) is at location 0. (It is assumed here that the storage locations are numbered 0 through n.)
- 10 When this pointer is at 0, the value of the other pointer is equal to the distance between the pointers. Thus, rather than actually having to calculate the spacing between the pointers, it is only necessary to read the value of one of the pointers. The determination of the pointer spacing when one of the
- 15 pointers is at 0 also provides a convenient timing mechanism. In other words, the pointer spacing is determined whenever the pointer is set to 0. If there are eight storage locations in the buffer, the spacing is determined every eighth read or every eighth write to the buffer. Because the difference
- 20 between the transmit and receive clocks is small, the spacing will change very slowly, so checking it only periodically is sufficient.

- 25 While the present invention has been described with reference to particular embodiments, it should be understood that the embodiments are illustrative and that the scope of the invention is not limited to these embodiments. Many variations, modifications, additions and improvements to the embodiments described above are possible. It is contemplated
- 30 that these variations, modifications, additions and improvements fall within the scope of the invention as detailed within the following claims.